DOCUMENT RESUME

ED 472 859 IR 058 565

AUTHOR	de Carvalho, Joaquim; Cordeiro, Maria Ines
TITLE	XML and Bibliographic Data: The TVS (Transport, Validation and Services) Model.
PUB DATE	2002-08-00
NOTE	14p.; In: Libraries for Life: Democracy, Diversity, Delivery. IFLA Council and General Conference: Conference Programme and Proceedings (68th, Glasgow, Scotland, August 18-24, 2002); see IR 058 549.
AVAILABLE FROM	For full text: http://www.ifla.org.
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	*Bibliographic Databases; *Bibliographic Records; Bibliographic Utilities; *Information Systems; *Library Services; Models; Online Systems; Programming Languages
IDENTIFIERS	*Bibliographic Services; MARC; *XML

ABSTRACT

This paper discusses the role of XML in library information systems at three major levels: as are presentation language that enables the transport of bibliographic data in a way that is technologically independent and universally understood across systems and domains; as a language that enables the specification of complex validation rules according to a particular data format such as MARC; and,finally, as a language that enables the description of services through which such data can be exploited in alternative modes that overcome the limitations of the classical client-server database services. The key point of this paper is that by specifying requirements for XML usage at these three levels, in an articulated but distinct way, a much needed clarification of this area can be achieved. The paper concludes by stressing the importance of advancing the use of XML in the real practice of bibliographic services, in order to improve the interoperable capabilities of existing bibliographic data assets and to advance the World Wide Web integration of bibliographic systems on a sound basis. (Contains 44 references.) (Author)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 472 859

IFLA GLASGOW 2002

68th IFLA Council and General Conference August 18-24, 2002

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

S. Koopman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Code Number: 075-095-E
Division Number: VI
Professional Group: Information Technology
Joint Meeting with: -
Meeting Number: 95
Simultaneous Interpretation: -

XML and bibliographic data: the TVS (Transport, Validation and Services) model

Joaquim de Carvalho
BookMARC; University of Coimbra
Portugal
E-mail: jrc@bookmarc.pt

Maria Inês Cordeiro
Art Library, Calouste Gulbenkian Foundation
Portugal
E-mail: ilopes@gulbenkian.pt

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Abstract:

*This paper discusses the role of XML in library information systems at three major levels: as a representation language that enables the **transport** of bibliographic data in a way that is technologically independent and universally understood across systems and domains; as a language that enables the specification of complex **validation** rules according to a particular data format such as MARC; and, finally, as a language that enables the description of **services** through which such data can be exploited in alternative modes that overcome the limitations of the classical client-server database services. The key point of this paper is that by specifying requirements for XML usage at these three levels, in an articulated but distinct way, a much needed clarification of this area can be achieved. The authors conclude by stressing the importance of advancing the use of XML in the real practice of bibliographic services, in order to improve the interoperable capabilities of existing bibliographic data assets and to advance the WWW integration of bibliographic systems on a sound basis.*

IR058565

BEST COPY AVAILABLE

1. Introduction

The application of XML (eXtensible Markup Language) in libraries has been drawing considerable interest, almost since it entered the realm of becoming a potential Web standard with universal impact [1]. Several exploratory projects are underway and some major libraries are committed to providing XML versions of their records. There is a mailing list on the topic and the literature on the subject is growing steadily [2].

Nevertheless, the widespread interest of the library community has not led to clearly perceived directions so far. Although there is consensus that XML should have some role in library information systems, it appears not clear what that role should be and how it will relate in practice to existing standards and systems, in practical terms [GARDNER; EXNER & TURNER,1998; SPERBERG-MCQUEEN, 1998; LYNCH, 2000; MCCALLUM,2000; MILLER, 2000, 2002; QIN, 2000; VAN HERWIJNEN,2000]. The most common topic of discussion is the relation of MARC [3] formats to XML, ranging from the current adequacy of MARC to what advantages could be gained by providing XML records for exchange rather than plain MARC files [HOPKINSON, 1998; LAM, 1998; MEDEIROS, 1999; GRANATA, 2000; MILLER, 2000, 2000a, 2002; JOHNSON, 2001]. More specific points of interest are issues about the "correct" form of XML representation for bibliographic records. In a brief overview we can identify the following patterns of usage:

1. Simple MARC to XML conversions: a direct equivalence between MARC elements and XML tags is proposed. The resulting XML DTD is relatively simple [4] and can be thought as equivalent in functionality to ISO 2709 [5];

2. Semantically rich MARC to XML conversions: an XML structure is designed that somehow reflects the meaning of the various Marc data elements. In some cases this conversion has the purpose of creating self-explainable (BiblioML) or self-validating (LoC) records [6]. This involves expressing every valid combination of fields and subfields into single XML elements. The resulting XML records are very complex and DTDs can run over 10.000 lines. The kind of difficulties of this approach is illustrated by [HOUGH, BULL & YOUNG,2000].

3. Simplified element sets for easy integration in external information systems: MARC is converted into a simplified structure which in turn is converted into XML. Reverse conversion is not possible due to significant loss of information [7];

4. XML conversion of Z39.50 [8] services.

In our view the current state of discussion could benefit from some conceptual clarification. The use of XML in libraries can span over a wide range of situations and purposes [HJØRGENSEN, 2001, 2001a]. We identify three major areas in which the adoption of XML could have significant impact: Transport, Validation and Services (hence the TVS acronym). Our argument is that by separating these areas one can more easily devise the targets for applying XML and make more clear particular arguments which, otherwise, would be difficult to address in general terms. We further propose specific requirements for a "normalized" XML usage in the areas concerned and put forward examples that cope with such requirements.

2. Why XML?

The importance of XML appears unquestionable simply by the fact that it is a language/format capable of representing complex structures in non-proprietary and self-explanatory ways. Although this seems enough, it is worth a closer insight to the understandings of what this assertion conveys.

"Universality" of the representation language and "enrichment" of represented contents seem to be the fundamental strengths of XML upon which the vision of the "semantic Web" has been built [BERNERS-LEE, HENDLER & LASSILA, 2001]. A vision where data on the Web is defined and expressed in such a way that enables better automated processing than HTML, not only for information presentation but also for purposes of functional integration, by allowing re-use of data by different software applications. Nevertheless, it became a common *cliché* to refer to the audience of the semantic Web as being "*not only* people but *also* computers". In our view, however, it is clear that the primary audience of the changes that led the so called "HTML cycle" [BERNERS-LEE, 1998] to a close, is computers, not people [BERNERS-LEE,1998a; W3 CONSORTIUM, 2001].

On the other hand, "richer content representation" to solve the problems of information discovery and retrieval of Web resources (meaning marked up documents) has been the most addressed aim of XML in the library and information environment, especially regarding pre-publishing metadata and metadata for digital libraries [ARMS, 2000; CAPLAN, 2000; DILLON, 2000; DOVEY, 2000; KIM & CHOI, 2000; QIN,2000].

Yet, XML also has a fundamental role with respect to the so called "hidden Web" [PRICE, 2002], meaning databases or other kinds of data assets that are out of the reach of Web search engines. Bibliographic databases are an important part of such hidden Web, included in what is often referred to as "legacy" systems. "Legacy systems" encompass both data assets and code that cannot be directly integrated in the Web technological environment due to the specificity of their models/formats, programming languages and platforms. Interoperability among them, and with the network environment in general, raises the need of an integrated view of issues concerning data access and re-use [DAY, 2000] and also of functional and affordable solutions to support it. On the whole, this means *simple protocols*, *clarified concepts* and *syntaxes*, *low barriers* to access and *tangible utility* of implementations [DUMBILL, 2001].

The aspects highlighted above are of utmost importance as a starting point for what this paper addresses. Therefore, our concern here is more focused on the potentials of XML to make independent applications talk to information object surrogates residing in whatever systems, irrespective of the applications they themselves run.

From this perspective, XML is important because it lowers the cost of developing applications that need to exchange complex data. Bibliographic data is complex as much as traditional data structures can be; thus its processing by computers will naturally benefit from the adoption of XML, especially when considering it in an open distributed environment.

Underpinning the cost-effectiveness of XML is the wide availability of software tools to develop XML-aware applications [9]. This meets the desired effects of the early standardisation of APIs (application programming interfaces) where realisation is now facilitated and improved with the definition of XML as a rich, yet general-purpose, notation language. In other words: XML is not just another notation, it is also a set of ready made and powerful tools that allow programmers to quickly implement complex data handling functionality in their applications.

Good XML usage should be driven by the cost/benefit factors brought in by such newly available tools, aiming at producing software components that are relevant to library practices. As stated above, people often forget that XML was created to be processed by computers, not by humans. The fact that an XML version of a bibliographic record is more machine-readable in an environment wider than that of ISO 2709, does not mean that a good criterion for XML representations of bibliographic data is to attain more clarity or simplification in record representation, and especially not for human consumption.

ISO 2709 is a transmission format designed to be handled efficiently by applications of a given domain, which have a long past and originate in a particular technological context, where library software needed to exchange data in sequential files, notably in magnetic tapes. In the same spirit, the use of XML should now be adjusted to and oriented by what are the most efficient data handling criteria and methods in today's technological context.

In order to analyse more clearly such criteria and methods we propose considering separately – although in an integrated manner – the use of XML in the areas of

I) ***record exchange*** (data transportation level),
II) ***record validation*** (data conformity level) and
III) ***sharing of services*** (application services level), not only among library information systems (with a role similar to that of Z39.50) but with a potentially very wide range of applications.

We believe that the introduction of XML can bring significant benefits to each of these areas especially if agreement could be attained in a set of higher level standards, which raises some organizational implications [SPERBERG-MCQUEEN, 1998; BRANDT, 2001].

3. Separating *transport* from *validation*

Some current proposals in the area of XML representation of bibliographic records try to achieve something that we believe is not a good target: an XML format specification that only allows valid bibliographic records to be represented. By "valid" we mean records that are fully conformant with the structure and rules defined in data formats such as MARC 21 or UNIMARC. In this perspective, a "good" XML format specification is one that makes impossible to represent an invalid record. This logic is behind the LoC XML specifications for bibliographic records and others inspired by it (see **Note 6**). Apparently, the improvement over formats like ISO 2709 seems enormous: while now we can have a valid ISO 2709 record that contains a totally invalid MARC record, in an XML representation the transport format would only allow semantically valid records to be exchanged.

In our view, this logic is based on a misconception about the role of XML and it leads to such complex XML format specifications that the cost/benefit of XML adoption easily evaporates. Furthermore, it fails to produce relevant functional benefit.

We believe that the misconception behind the idea of format specifications that enforce validation arises from similarities between the kind of rules used to validate XML documents and the kind of rules that define data formats like MARC21 or UNIMARC. Both XML documents and MARC records are tree-like structures, composed of elements that can repeat themselves and further be divided into other elements (nodes in XML, fields, indicators, subfields in MARC).

In the XML world, the rules that a well-formed piece of information must conform to can be defined formally in a special document called a DTD (Document Type Definition) and, more recently, through XML documents called *Schema* [10]. Schema and DTDs allow programmers to define a valid XML

structure by specifying which are the elements it can contain, in which order they can appear, what is mandatory and/or repeatable, etc... By providing this definition the tools available for XML processing can automatically detect invalid documents.

There is in fact a remarkable parallelism between the rules that can be written in DTDs and Schema and the rules that define a valid MARC record. In both cases we have assertions about repeatability of data elements and subelements and rules for when they are mandatory or optional, etc. But on closer inspection we see that the semantics of MARC are much more complex than those that can be defined for valid XML. Besides, the elementary structural components of MARC records are also more complex that those of XML: in the latter we have *Attributes* and *Elements* and in the former we have *Leader*, *Fields*, *Indicators* and *Subfields*, and even data offsets/limited length for certain coded data subelements. On top of this, MARC also encompasses application rules that include a high number of conditions that rely on human discrimination, but that affect the more or less validity of the record (for example, while Field X and Field Y are both valid they should never appear together in the same record if the record contains Field or Subfield Z... or a given value in Field W...).

To conclude: we have a fairly simpler formalism trying to grasp a far more complex one. To achieve the goal of an auto-validating XML/MARC format specification it is necessary to describe all MARC records possibilities into artificially complex XML structures, designed with the sole purpose of making the DTDs and SCHEMA validation rules to express MARC syntax/semantics.

In other words, creating an XML representation of MARC records with the goal of making it self-validating produces extremely complex XML records. The resulting XML structure has no strait equivalence to the MARC structure (Record, Fields, Subfields). Many intermediate levels are introduced on the XML side to allow the crude rules of DTDs to enforce the complex rules of MARC.

We see no functional benefit in this. From an application design point of view, semantic errors in the incoming data cannot efficiently be treated just as format (syntactic errors). An application will always need an high level of internal representation of the MARC semantics that can be applied in modules other than record transportation. Even the records that pass through as valid will have such a complex structure that rebuilding the MARC information from them is no trivial task.

Instead of trying to merge structure, syntax and semantics in a single XML format specification we propose two separate standards to emerge:

1. The specification of an XML/MARC transportation format, equivalent in requirements and functionality to the role of the current ISO 2709. Its purpose is to allow the efficient transport of bibliographic data. Like ISO 2709 this format contains the necessary information for representing the morphological structure of the MARC record, but does not aim to attain validation of the complete syntax/semantics of the MARC format. This format would map directly to the MARC record main structural levels. Our examples of such a specification can be found at http://www.bookmarc.pt/tvs.

2. A specification of an XML format for expressing MARC-based semantics. This can be thought of as an XML representation of the MARC Manual. This allows the full expression of all syntax/semantic rules of MARC records, including all context explanations, examples, etc, expressed in a machine-readable format, usable either by humans and machines. A full version of the UNIMARC Manual in XML was developed with this aim and is available for demonstrations at http://www.bookmarc.pt/tvs.

BEST COPY AVAILABLE

Both specifications 1 and 2 are designed in a coordinated matter, so that the validation of a record could be done by applying the machine-readable rules to a record represented in XML transportation format.

4. Introducing Web services and WSDL

In designing requirements for the above mentioned standards careful consideration should be given to the emergence of the Web Services paradigm as the major interoperability model in the computer industry [11]. Web services paradigm provides a model for inter-application communication based on easy to implement standards. The main components of this paradigm are:

- HTTP as transport protocol;
- XML as a data exchange format language;
- SOAP [12] as a message structure format;
- WSDL (Web Services Description Language) [13] as a tool for self description of available services;
- UDDI (Universal Description, Discovery and Integration) [14] and ebXML [15] as standards for directories of available services.

Basically, Web services are remote services provided by information systems to remote applications, over HTTP, exchanging data in SOAP messages. The services auto-describe themselves through WSDL and directories are available for automatic browsing so that every provider for a given service can be located.

This model targets the same problem domain as Z39.50, but in a broader context that aims at universal application. There is, however, a huge difference in technical and industrial context: Web services are being pushed with very significant resources by Microsoft, Sun, IBM and many other important corporations of the software industry. The level of ubiquity that can be achieved by a library service that is made available as a Web service is immensely greater than what can be attained through Z39.50.

Apart from these technical reasons, there is also – as pointed out for XML – cost related reasons that favor Web services. Web services are self-describing (using WSDL) in a way that the needed code for a client to access a given service can be automatically generated. This means that a software client for a catalog search can be created in a few minutes, by pointing a software development tool like *Microsoft Visual Studio*™ to the URL of a library [16]. Through directory services based on UDDI and ebXML it is possible to automate, almost completely, the generation of client software that provides a very high level access to information services.

5. Moving service descriptions to WSDL

However, the automation achieved through WSDL will be limited if the semantics of the records and the details of the service interfaces vary widely among institutions. At this level the problems that have limited Z39.50 usefulness [HINNEBUSH, 1998; LE VAN, 1998; GATENBY, 2000; MOEN, 2000, 2001] will not disappear just because the base technology of interoperability has changed. An extra level of standardization would be necessary. The target here is to agree in a set of common operations, eventually build from the Z39.50 services. Some experiments and prospects about using XML and Web services in Z39.50 are now taking shape (see **Note 9**), mostly by members of the ZIG, since 2000 [ZIG; ZAGALO, MARTINS & PINTO, 2001].

Taking the original purpose of Z39.50 to the new paradigm would result in a set of service descriptions, expressed in WSDL. Software vendors would implement Web services for these descriptions. Currently there are tools available that take a WSDL description of a service and create a token server component,

which deals with all the details of communicating with clients, leaving the implementer with the only task of actually bridging the generated code to a specific library system. This may appear to be what the Z39.50 libraries have been doing all along. In theory that is true, but in practice the level of integration with modern development environments is vastly different. Again, the cost factor is very relevant, especially because in the very near future every professional computer programmer will have to have all the skills needed to deal with web services technologies and this is not, and will not be, the case with Z39.50.

The tasks of achieving some level of standardization around library services should not prove more difficult in the SOAP-WSDL environment than it was previously. The long experience with Z39.50 and ILL protocols is itself a solid basis on which to start.

In our view, such standardization efforts are needed and should provide the necessary specifications for XML formats for bibliographic information. In the context of WSDL, descriptions of complex data formats in XML are constrained by the necessity of providing automatic serialization and de-serialization of information to and from data types in modern object-oriented languages such as C# and Java. In other words, the major requirement that an XML format for MARC records faces in today's technological context is to provide automatic de-serialization into C# and Java, or any other object-oriented language (the more the better) so that the automatic generation of client and server code from WSDL descriptions can work transparently for software developers.

These aspects may seem a set of very fine technical points but the fact is that they are the cornerstones of our argument in this paper. XML can take the role of ISO 2709 for the Web environment, and be the basis of the distributed library services of the future. While the main requirements behind ISO 2709 were to achieve the best integration with the technological tools available at the time, the same requirements, when applied to a future XML standard for bibliographic data exchange should produce something that transparently maps the complex bibliographic data types within the framework of automatic code generation from WSDL descriptions. This completely rules out the complex formats referred to early on and points rather to more direct mappings like OAI's.

We strongly believe that the requirements for XML as a transport format for bibliographic data should be mostly derived from the future envisaged for XML as the basis of a new generation of interoperable services. Attempts to contaminate XML record formats with semantic requirements, like self-validation or human readability, will impair its ability to fulfill this central role. Metadata for semantics and validation of XML MARC records should be formalized through separate standards, as argued above [17].

6. Conclusions

The proposals advanced in this paper assume that the future of library services will strongly rely on the adoption of Web technology, at a structural level that goes far deeper than just providing web interfaces to library systems. In devising such a future one central aspect to consider is the model of interoperability that will enable us to reinforce and diversify the role of libraries, and the usefulness of their structured information assets. This is especially important for the provision of information services aimed at being more ubiquitous and re-usable, within and beyond the library community. This understanding encompasses two major foci of attention: the language that libraries use to talk to their environment, concealing all data representational issues, and the services libraries provide in the open network.

The first focus involves a special attention to all aspects concerning the impact of XML technology in the management of bibliographic data, namely in terms of transportability and re-usability. This has provided opportunity to revisit the place and function of some of the library data standards in the light of the new Web language: why, how, and what for should libraries talk XML? So far, the answers to these questions

are still diffuse and do not show clear directions. Our main argument regarding the current state of affairs in this matter is that the most important now is to clarify targets and methodologies, prior to invest in the development of actual solutions. At this level we have proposed a methodology that relies on the separation of different concerns: data transport, data validation and services. We explained why the issues raised in each of these levels are better addressed separately and how such an approach increases the efficacy and cost effectiveness of actual solutions.

The second main focus are services – mostly concerned with the "what for" side of changes brought about by Web technology. Following all the developments around XML and related XML constructs in the last four years, Web services are currently the latest buzzword in the field. It is now clear from all the industry and standard stakeholders activities that Web services will be a foundation for dramatically changing the e-business environment, offering cost effective solutions for what has been for so long the most demanding and expensive side of applied IT – interoperability at the application level for inter-systems transactions.

The library community was among the earliest professional communities to face the need for standards to support interoperability at the application level, therefore benefiting from long experience and deep knowledge about the multiple levels of requirements involved. Namely the perception that most of such requirements go back to the level of data standards quality. We, therefore, called attention to the need for standards concerning the use of XML to handle bibliographic data, with the understanding that XML alone only provides the language, not the dictionary, syntax and semantics that should be known to support XML conversations in a particular domain. Furthermore, we put forward the proposition that such standards should be agreed separately for transport and validation, arguing that this separation may strongly facilitate the implementation of Web services in the library environment. Such standard efforts are essential to make the most of all the benefits that Web services technology has to offer to libraries - benefits which are already being recognized, *e.g.*, in the directions taken by the Z39.50 research and development groups.

It may sound like the main debate regarding these matters is of IT technicalities. In our view it is much more than that. Although tiny technicalities populate these topics, we should not let the details to cloud our perception of the big picture. The big picture is where we can change the business principles, targets and rules. The opportunities that Web technology is bringing in can reshape substantially the "library business", depending on whether or not we pick the right vision and adopt strategies accordingly.

Examples and demonstrations

More details on the **TVS Model**, including XML examples for the different levels proposed, and practical demonstrations of the bibliographic Web services created, can be found at: http://www.bookmarc.pt/tvs.

Notes *(All URLs valid as of April 15 2002)*

[1] First issued in 1998, with the status of a W3 Consortium Recommendation, XML is only the central piece of a much larger set of specifications, on the whole usually referred to as XML technology. In this acceptation, XML is a rapidly expanding technology with an exponentially growing literature. The central reference point is the W3 Consortium: http://www.w3.org/XML/.

[2] The mailing list is *XML4LIB Electronic Discussion* and provides a good forum for exchanging ideas and keeping up with current trends (http://sunsite.berkeley.edu/XML4Lib/). Some sources for bibliography on the topic are *The XML Cover Pages: MARC (MAchine Readable Cataloging) and SGML/XML*, by Robin Cover (http://www.oasis-open.org/cover/marc.html); *A Webliography of XML and XML/MARC Related Links*

9
BEST COPY AVAILABLE

(http://xmlmarc.stanford.edu/webliography.html) as well as *Medlane XMLMARC*, by Dick Miller (http://xmlmarc.stanford.edu/).

[3] MARC – *Machine Readable Cataloging* – refers generally to library data formats, which may encompass different subsets for bibliographic, authority, holdings, classification and community information; in this paper we refer to MARC meaning mainly bibliographic data formats, of which the most representative at an international level are currently MARC21 (http://lcweb.loc.gov/marc/) and UNIMARC (http://www.ifla.org/VI/3/p1996-1/sec-uni.htm).

[4] The best example is the MARC XML format of the *OAi- Open Archives Initiative* (http://www.openarchives.org/OAI/oai_marc.xsd); Java source code for handling the format is available (http://www.dlib.vt.edu/projects/OAi/marcxml/marcxml.html). Another example of a direct mapping DTD, also with Java source code, is included in *JAMES -Java Marc EventS* - by Bas Peters (http://www.bpeters.com). *MARC.pm*, another software package, in Perl, also introduces a similar format (http://marcpm.sourceforge.net/) . All these formats have in common the fact of being designed for easy processing by computer programs.

[5] ISO 2709:1996 - *Information and documentation -- Format for Information Exchange.* ISO 2709 is the common format for exchange underlying all MARC formats; it consists of a record label, a directory and data fields, with standard characters for separators. (http://www.iso.org/)

[6] For the LoC - *Library of Congress* - MARC DTDs philosophy see: http://lcweb.loc.gov/marc/marcdtd/marcdtdback.html and for examples, files and utilities: http://lcweb.loc.gov/marc/marcsgml.html . The *Danish Bibliographic Centre* (http://www.dbc.dk) and *Portia Systems* (http://www.portia.dk) were producing DTDs for DANMARC and UNIMARC similar to that of LoC, according to a message from Poul Henrik Jørgensen of Portia to the XML4LIB mailing list, May 14th 2001 (see http://sunsite.berkeley.edu/XML4Lib/archive/0105/0023.html). *BiblioML*, sponsored by the French Ministry of Culture, converts UNIMARC fields and subfields into XML elements, whose names reflect their actual meaning masking the tag/subfield identifiers of the original record (http://www.culture.fr/BiblioML). Another example of rich conversion is the *Medlane XMLMARC* experiment (http://xmlmarc.stanford.edu).

[7] The *Dublin Core Metadata Element Set – DCMES* is the most commonly adopted standard for simplified XML conversion (http://dublincore.org/documents/2001/04/11/dcmes-xml/). It is adopted in the *Z39.50 Bath Profile* (http://www.nlc-bnc.ca/bath/bath-e.htm) as the record syntax for *Functional Area C - Cross-domain search and retrieval for resource discovery* (http://www.nlc-bnc.ca/bath/bp-app-d.htm). Recently the Library of Congress started an effort to define an XML schema for MARC21 that would provide an encoding richer than Dublin Core: *MODS - Metadata Object Description Schema* (http://www.loc.gov/standards/mods/).

[8] Z39.50 is used to refer either to *ANSI/NISO Z39.50* or to *ISO 23950* standards, as they parallel the same specifications for the protocol.
ANSI/NISO Z39.50.. Information Retrieval. Application Service Definition and Protocol Specification For Open Systems Interconnection. (http://www.niso.org/standards/resources/Z3950_Resources.html).
ISO 23950:1998. *Information and Documentation. Information Retrieval (Z39.50). Application Service Definition and Protocol Specification.* (http://www.iso.org/)
The Library of Congress is the *International Standard Maintenance Agency* for Z39.50 (http://www.loc.gov/z3950/agency/); the *ZIG - Z39.50 Implementers Group* works for the developments of the standard (http://lcweb.loc.gov/z3950/agency/zig/zig-meetings.html).
Current Z39.50 development efforts seem to converge on the need to adapt the standard to more modern architectures, like Web services, where XML plays a central role at various levels. An initiative is underway called *ZING - Z39.50 International Next Generation* (http://www.loc.gov/z3950/agency/zing/zing.html). The impact of Web services on the future of XML in libraries is discussed in the last part of the present paper.

[9] See, for example, the resource page *XML database products*, maintained by Ronald Bourret, at http://www.rpbourret.com/xml/XMLDatabaseProds.htm.

[10] The main difference between a DTD and an XML Schema is that the latter uses XML syntax, therefore it does not imply knowledge of any other syntax and will have the same advantages as XML such as extensibility. XML Schema defines how to express structure, attributes, data-typing, etc. of XML documents. *XML Schema* is a W3C Recommendation since May 2001, available at http://www.w3.org/TR/xmlschema-0/.

[11] An overview of Web services is provided by the *Diffuse Project*, sponsored by the European Commision's IST Programme (http://www.diffuse.org/WebServices.html). Web services are a central part of *Microsoft .NET* initiative (http://www.microsoft.com/net/default.asp), which is gaining great impact on software development methodologies. Web services is a topic that begins to appear in library-oriented literature [BURRIDGE, 2001; GARDNER, 2001] but few library projects involving Web services are known so far.

[12] *SOAP* acronym stood originally for *Simple Object Access Protocol*, but the word is now used as the full name of the protocol, since the last revision of the standard. Basically SOAP is a format for requesting services from remote applications through XML. SOAP is a central part of Web services . Current status of the standard can be found at http://www.w3.org/2002/ws/.

[13] *WSDL - Web Services Description Language*. The idea behind WSDL is to make web services self documented for client applications. WSDL is target at software development tools, not at humans. A tool should be able to read a WSDL description of a service and automatically produce most of the client application code needed to access the described service. See: http://www.w3.org/TR/wsdl.

[14] *UDDI* addresses the need for international registries of Web services; it encompasses both the specifications needed to register a service and actual registry services as well. Directories of Web services are a central component of the new paradigm. See http://www.uddi.org/.

[15] *ebXML* (*Electronic Business using eXtensible Markup Language*) is an initiative sponsored by the *UN/CEFACT* and *OASIS* (http://www.ebxml.org/) aimed at making *EDI* (Electronic Data Interchange) semantics available in XML. It consists of a modular suite of specifications especially designed for electronic commerce Web services, including directory of services similar to UDDI. The *Diffuse Project* offers an overview of the ebXML specifications (http://www.diffuse.org/ebXML.html#RR).

[16] For an explanation about how to quickly build client code for a library search Web service see our documentation at BookMARC: http://ptolemy.bookmarc.pt:8001/kb/kb000002.html.

[17] Examples are provided in the *TVS Model* Web page at http://www.bookmarc.pt/tvs/.

References *(All URLs valid as of April 15 2002)*

1. ARMS, C. (2000) Some observations on metadata and digital libraries. *Library of Congress Bicentennial Conf. on Bibliographic Control for the New Millenium, Washington, 2000* [online]. Available at: http://lcweb.loc.gov/catdir/bibcontrol/arms_paper.html.

2. BERNERS-LEE, T. (1998) *Evolvability* [online, up. 2000]. Available at: http://www.w3.org/DesignIssues/Evolution.html.

3. BERNERS-LEE, T. (1998a) *Web architecture from 50,000 feet* [online, up. 2002]. Available at: http://www.w3.org/DesignIssues/Architecture.html.

4. BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. (2001). The semantic web. *Scientific American*, May (2001) [online]. Available at : http://www.sciam.com/2001/0501issue/0501berners-lee.html.

5. BRANDT, D.S. (2001) Clarifying XML. *Library Technology*, May (2001) [online]. Available at: http://www.emeraldinsight.com/librarylink/technology/may01.htm.

6. BURRIDGE, B. (2001) Windows NT Explorer: Simple Object Access Protocol (SOAP). *Ariadne*, no 29 (Sept. 2001) [online]. Available at: http://www.ariadne.ac.uk/issue29/ntexplorer/.

7. CAPLAN, P. (2000) International metadata initiatives: lessons in bibliographic control. *Library of Congress Bicentennia. Conf. on Bibliographic Control for the New Millenium, 2000* [online]. Available at: http://www.loc.gov/catdir/bibcontrol/caplan_paper.html.

8. DAY, M. (2000) Resource discovery, interoperability and digital preservation: some aspects of current metadata research and development. *VINE*, 117, pp. 35-48.

9. DILLON, M. (2000) Metadata for Web resources: how metadata works on the Web. *Library of Congress Bicentennial Conf. on Bibliographic Control for the New Millemium, 2000* [online]. Available at URL: http://www.loc.gov/catdir/bibcontrol/dillon.html .

10. DOVEY, M. J. (2000) "Stuff" about "Stuff": the different meanings of metadata. *VINE*, 116 (2000), pp. 6-13.

11. DUMBILL, E. (2001) Building the semantic Web. *XML.com*, 2001.03.07 [online]. Available at: http://www.xml.com/pub/a/2001/03/07/buildingsw.html.

12. EMMOTT, S. (1998) At the event: SGML, XML and databases. *Ariadne*, 18 (Dec. 1998) [online] Available at (http://www.ariadne.ac.uk/issue18/emmott-sgml/.

13. EXNER, N.; TURNER, L. (1998) Examining XML: new concepts and possibilities in Web authoring. *Computers in Libraries*, vol. 18, no 10 (Nov./Dec. 1998) [online]. Available at: http://www.infotoday.com/cilmag/nov98/story2.htm .

14. GARDNER, J. R. [undated] *Exploring What's neXt: XML, information sciences and markup technology* [online]. Available at: http://vedavid.org/xml/docs/eXploring_xmlandlibraries.html.

15. GARDNER, T. (2001) An introduction to Web services. *Ariadne*, 29 (Oct. 2001) [online]. Available at: http://www.ariadne.ac.uk/issue29/gardner/intro.html .

16. GATENBY, J. (2000) *Internet, interoperability and standards: filling the gaps* [online]. Available at http://www.niso.org/press/whitepapers/Gatenby.html.

17. GRANATA, G. (2000) XML e formati bibliografici. *Bolletino AIB*, no 2 (2000) pp.181-191.

18. HINNEBUSH, M. (1998) Z39.50: *Report to the CIC on the state of Z39.50 within the Consortium* [online]. Available at : http://www.cic.uiuc.edu/cli/Z3950/z39-50report.htm.

19. HJØRGENSEN, P. H. (2001*) XML standards and library applications*. Paper presented at *ELAG 2001*.[online]. Available at: http://www.stk.cz/elag2001/Papers/Poul_HenrikJoergensen/Show.html.

BEST COPY AVAILABLE

20. HJØRGENSEN, P. H. (2001a) *VisualCat: cataloging with XML, RDF, FEBR & Z39.50* [online]. Available at: http://www.bokis.is/iod2001/slides/Jorgensen_slides.ppt.

21. HOPKINSON, A. (1998) Traditional communication formats: MARC is far from dead. *International Seminar "The Function of Bibliographic Control in the Global Information Infrastructure"*[online]. Available at http://www.lnb.lt/events/ifla/hopkinson.html.

22. HOUGH,J.; BULL, R.; YOUNG, B. (2000) *Using XSLT for XML MARC record conversion. Discussion paper, v. 0.2*, 16.06.2000 [online]. Available at: http://roadrunner.crxnet.com/one2/xslt_marc_report.pdf .

23. KIM, H.; CHOI, C. (2000) XML: how it will be applied to digital library systems. *The Electronic Library*, vol. 18, no 3 (2000) p. 183.189.

24. JOHNSON, B. C. (2001) XML and MARC: which is right? *Cataloging and Classification Quarterly*, vol.32, no 1 (2001) pp. 81-90. Also available online at: http://elane.stanford.edu/docs/johnson.pdf.

25. LAGOZE, C.; VAN SOMPEL, H. (2001) *The Open Archives Initiative: building a low cost interoperability framework*. Paper presented at the *Joint Conference on Digital Libraries*, Roanoke VA, 2001 [online]. Available at: http://www.openarchives.org/documents/oai.pdf.

26. LAM, K. T. (1998) *Moving from MARC to XML* [online]. Available at: http://ihome.ust.hk/~lblkt/xml/marc2xml.html.

27. LE VAN, R. (1998) *Library experience in online searching: a position paper for the Query Languages Worksop, 1998* [online]. Available at: http://www.w3.org/TandS/QL/QL98/pp/libraryexp.html.

28. MCCALLUM, S. (2000) Extending MARC for bibliographic control in the Web environment: challenges and alternatives. *Library of Congress Bicentennial Conf. on Bibliographic Control for the New Millemium*.[online] Available at: http://lcweb.loc.gov/catdir/bibcontrol/mccallum_paper.html

29. MEDEIROS, N. (1999) Making room for MARC in a Dublin Core World. *Online*, Nov. 1999 [online]. Available at: http://www.onlinemag.net/OL1999/medeiros11.html.

30. MILLER, D. R. (2000) XML: libraries' strategic opportunity. *Library Journal. NetConnect, Summer 2000* [online]. Available at: http://www.libraryjournal.com/xml.asp .

31. MILLER, D. R. (2000a) *XML and MARC: a choice or a replacement?* [online]. Paper presented at *ALA Annual Conf, 2000*. Available at: http://elane.stanford.edu/laneauth/ALAChicago2000.html.

32. MILLER, D. R. (2002) *Adding luster to librarianship: XML as an enabling technology* [online]. Available at: http://elane.stanford.edu/laneauth/Luster.html.

33. MOEN, W. (2000) Resource discovery using Z39.50: promise and reality. *Library of Congress Bicentennial Conf. on Bibliographic Control for the New Millemium* [online]. Available at: http://lcweb.loc.gov/catdir/bibcontrol/moen.html.

34. MOEN, W. E. (2001) *Improving Z39.50 interoperability: z39.50 profiles and testbeds for library applications*. Paper presented at the *67th IFLA Council and General Conference, 2001* [online]. Available at: http://www.ifla.org/IV/ifla67/papers/050-203e.pdf.

BEST COPY AVAILABLE

35. MOTTA, S.; URSINO, G. (2000) XML su tecnologia MOM: un nuovo approccio per i software delle biblioteche. *Bolletino AIB*, no 2 (2000) pp.195-203.

36. PREMINGER, M.; HOLM, L. (1997) *MARC conversion: a practical approach* [online]. Paper presented to the *Library Systems Seminar, Gdansk, Poland, June 1997. Available in the* ELAG Archive, at: http://www.kb.nl/coop/elag/elag97/papers/marc_art.htm.

37. PRICE, G. (2002) The invisible Web [online]. Paper presented to the *Internet Librarian International 2002*, London. Available at: http://www.freepint.com/gary/ili2002.htm.

38. QIN, J. (2000) Representation and organization of information in the Web space: from MARC to XML. *Informing Science*, vol. 3, no 2 (2000) [online]. Available at: http://inform.nu/Articles/Vol3/v3n2p83-88.pdf .

39. SPERBERG-MCQUEEN, S. (1998) *XML and what it will mean for libraries*. [online] Available at http://tigger.uic.edu/~cmsmcq/talks/teidlf1.html.

40. VAN DE SOMPEL, H.; LAGOZE, C., *eds*. (2001) *The Open Archives Initiative Protocol for Metadata Harvesting. Version 1.1, July 2001* [online]. Available at: http://www.openarchives.org/OAI_protocol/openarchivesprotocol.html.

41. VAN HERWIJNEN, E. (2000) *The impact of XML on Library procedures and services* [online]. Available at: http://lhcb.web.cern.ch/lhcb/~evh/xmlandlibrary.htm.

42. W3 CONSORTIUM (2001) *Semantic Web Activity Statement* [online]. Available at: http://www.w3.org/2001/sw/Activity.

43. ZAGALO, H.T; MARTINS, J.A.; PINTO, J.S. (2001) Design and development of a virtual library and a SOAP/Z39.50 gateway using Java technologies. *Proc. SPIE*, vol. 4521, p. 52-61. Paper can be requested at SPIE Web (http://spie.org/).

44. ZIG *Z30.50 Implementors Group (ZIG) Meetings* [online]. Available at: http://lcweb.loc.gov/z3950/agency/zig/zig-meetings.html.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

X	This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.
	This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").